

# Incremental Learning Method for Wind Turbine Fault Detection Models Considering False Negatives

Ziqi Wang, Xiaohang Jin, *Senior Member, IEEE*, and Zhengguo Xu, *Member, IEEE*

**Abstract**—Fault detection (FD) algorithms based on supervisory control and data acquisition (SCADA) data have been widely used in the operation and maintenance of wind turbines (WT). However, the performance of FD models will degrade due to the time-varying operating conditions (TVOC). Incremental learning (IL) methods can be employed to update the models online to adapt to TVOC, but the error accumulation caused by false negatives will lead to a continuous decrease in the fault detection rate (FDR). A novel IL method considering false negatives is proposed to improve the performance of WT FD models. Firstly, a data buffer is built to cache some new normal data used in model updating. Secondly, a processing strategy for false negatives is proposed to block some high-risk data from being added to the buffer, thereby weakening the error accumulation. Thirdly, when the buffer is full, update the FD model using the data in it. A real-world SCADA dataset with gearbox faults and four different FD algorithms with various model updating strategies are used in the experiments. The results demonstrate that the proposed method can lower the false alarm rate (FAR) of all four FD algorithms. After processing false negatives, for example, for the FD model based on Gaussian kernel regression, its FDR increased by about 11%, and for the FD model based on multivariate state estimation technique, its FDR increased by about 14%. The results of hyperparameter experiments and early stopping experiments show that the proposed method has good potential for practical applications.

**Index Terms**—Wind turbine (WT), Fault detection (FD), Incremental learning (IL), Supervisory control and data acquisition (SCADA), Error accumulation.

## I. INTRODUCTION

The demand for new technologies in the operation and maintenance (O&M) of wind turbines (WT) has grown in recent years. It is crucial to monitor the conditions of WT by analyzing the operational data to detect incipient faults and reduce O&M costs [1]. Methods based on

supervisory control and data acquisition (SCADA) data have received a lot of attention for condition monitoring of WT [2–4]. However, there is an imbalance problem with SCADA data [5], meaning that there is much less fault data than normal data. The reasons for this include WT not being allowed to operate for a long time under fault conditions [6], inefficient manual labeling, etc. Therefore, fault detection (FD) algorithms based on normal behavior modeling [7, 8] have been widely used. These FD models are trained using only normal data and can detect faults by analyzing the differences between current and normal conditions.

Due to external factors such as changes in weather and season, as well as internal factors such as system settings adjustments and equipment retrofits, WT is often under time-varying operating conditions (TVOC) [1]. For example, the normal temperature of the generator bearing is higher in summer than in winter. And when the settings of the cooling system are adjusted, the normal range of the gearbox oil temperature will change. The accuracy of FD models will degrade under TVOC because training and test data are non-identically distributed, which originates from the complex, open, and dynamic operating environment of WT. Therefore, the training data of FD models should cover all possible normal conditions, which is difficult in practical practice, especially for new wind farms [9].

In addition to general strategies like regularly retraining the models, existing research has proposed some methods to improve the performance of WT FD models under TVOC. The dynamic threshold methods are used to enable the models to tolerate the parameter fluctuations caused by TVOC [10, 11], thereby reducing the false alarm rate (FAR). The transfer learning methods are used to make the data from other turbines applicable for training the FD model of the target WT [12, 13]. The covariate adjustment methods are used to decouple some temperature parameters from environmental factors [14, 15], which can improve the stationarity of these parameters under TVOC. However, the above methods may not be able to adapt to new types of TVOC in the data stream, resulting in the degradation of FD model performance during long-term operation.

Incremental learning (IL) refers to a learning paradigm that allows models to continuously learn new knowledge from new data while retaining the learned knowledge [16]. Using IL methods, FD models can continuously learn new knowledge online from the data stream, thereby adapting shortly to TVOC and maintaining good performance. In [17], a novel incremental support vector data description method was proposed for damage detection of blades based on acoustic

Manuscript received ; revised ; accepted . Date of publication ; date of current version . This work was supported in part by the Huzhou Natural Science Foundation under Grant 2023YZ14, in part by the National Key Research and Development Program of China under Grant 2022YFE0198900, in part by the National Natural Science Foundation of China (NSFC) under Grant 61973269, and in part by the Ningbo Natural Science Foundation under Grant 2021J038. (*Corresponding author: Zhengguo Xu.*)

Ziqi Wang and Zhengguo Xu are with the College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China, and also with the Huzhou Institute, Zhejiang University, Huzhou 313000, China (e-mail: wangziqincepu@163.com; xzg@zju.edu.cn).

Xiaohang Jin is with the College of Mechanical Engineering and the Key Laboratory of Special Purpose Equipment and Advanced Processing Technology, Ministry of Education and Zhejiang Province, Zhejiang University of Technology, Hangzhou 310023, China, and also with the Ninghai ZJUT Academy of Science and Technology, Ninghai 315600, China (e-mail: xhjin@zjut.edu.cn).

signals, and the model can be adaptively updated to improve accuracy. In [18], a framework based on deep learning and IL was proposed for fault identification of drive bearings using vibration signals, and it enables adaptive creation and modification of fault patterns. In [19], an FD method based on non-parametric regression and IL was proposed for the gearboxes using SCADA data, and the FAR is lower than the traditional models during long-term operation. IL methods for data streams include two specific strategies: (1) how to select specific data from the data stream for model updating, and (2) how to adjust parameters in the model. Existing research mainly focuses on the latter strategies, including the elastic weight consolidation strategy [20], dynamic structure strategy [21], replay strategy [22], etc. It should be noted that these strategies are only applicable to specific learning algorithms. In this paper, we focus on the strategies that select specific data to be used in model updating, which can be applicable to different FD algorithms and model updating strategies.

Besides the catastrophic forgetting problem where the learned knowledge is abruptly forgotten as new information is learned [23], models with IL also face the error accumulation problem. This means that errors will propagate from previous data to later ones through model updating, causing more errors [24]. For the FD models based on normal behavior modeling, only normal data is used in online updating, just like in offline training. The error accumulation of these FD models is caused by false negatives, which are defined as fault samples being misclassified as normal and also known as missing alarms. This leads to more false negatives and a continuous decrease in the fault detection rate (FDR) during long-term operation. It is common for certain parameters of WT to undergo short-term fluctuations due to TVOC and noises [25], which can result in significant false negatives and substantial error accumulation. To address this issue, an adaptive condition monitoring method for WT, based on non-parametric regression and continual learning, was proposed in [26]. In this method, a punishment mechanism is proposed for fault data and their adjacent data to improve the FDR. However, the punishment mechanism lacks flexibility, limiting its suitability only to high-frequency updating.

In this paper, a novel IL method for WT FD models is proposed considering the false negatives, and the proposed method is compatible with various FD algorithms. In the proposed IL method, normal data detected by the FD model in the data stream are cached in a data buffer. When the buffer is full, the FD model is updated using the data in it. To address the error accumulation, high-risk data that are adjacent to the detected fault data are identified as false negatives and prevented from being used in model updating. Specifically, when the FD model detects a fault data, part of the data previously cached in the buffer is deleted, and part of the subsequent data is blocked from being added to the buffer. A real-world WT SCADA dataset with gearbox faults is used to verify the proposed IL method through various experiments,

and the performance of the proposed method is discussed in four different FD algorithms.

The contributions of this paper are as follows:

1) A data buffer is built to cache some new data used in model updating. Normal data detected by the FD model in the data stream will be cached in a buffer, and the data in it will be used for the model updating when the buffer is full. The proposed method can be applied to different FD algorithms and is suitable for practical applications.

2) A processing strategy for false negatives is proposed to alleviate the error accumulation in FD models with IL. Considering the strong temporal correlation in the data stream, some high-risk data that are adjacent to fault data are identified as false negatives and they are blocked from model updating. The proposed strategy can improve the FDR of FD models during online updating.

The rest of the paper is organized as follows: Section II presents the framework of FD models with IL. Section III analyzes the error accumulation caused by false negatives and presents the proposed IL method in detail. Section IV presents the SCADA dataset and the processing of the dataset. Around experiments and related analyses are listed in Section V. Section VI discusses the conclusions.

## II. FRAMEWORK OF WIND TURBINE FAULT DETECTION WITH INCREMENTAL LEARNING

Due to the data imbalance problem, many FD algorithms using SCADA data are based on the framework of normal behavior modeling to train models and detect faults. And considering improving accuracy through model updating, IL methods can be used. The framework of WT FD models with IL based on normal behavior modeling is given in Fig. 1. This framework consists of the following three parts.

1) Offline modeling. Firstly, historical SCADA data collected under normal conditions is pre-processed, and two types of data should be processed. One type is data collected during shutdown, power limitation, and other similar conditions because the WT cannot output electricity normally during these times. Another type is data that contains gross errors, null values, or null timestamps caused by noise, transmission errors, etc. After pre-processing, the majority of the remaining historical data is normal data. Secondly, the normal data constitutes a training set, and the FD model is trained based on this training set and used in Online detection.

2) Online detection. Firstly, the SCADA data stream is pre-processed using the same rules as Offline modeling to prevent interference. For example, data collected under power limitation conditions may be misjudged as potential faults. Secondly, the processed real-time data is input into the FD model and the FD index is calculated. Thirdly, the FD result is obtained based on the FD index and the FD threshold. Generally, data above the threshold is classified as fault data, while data below the threshold is classified as normal data.

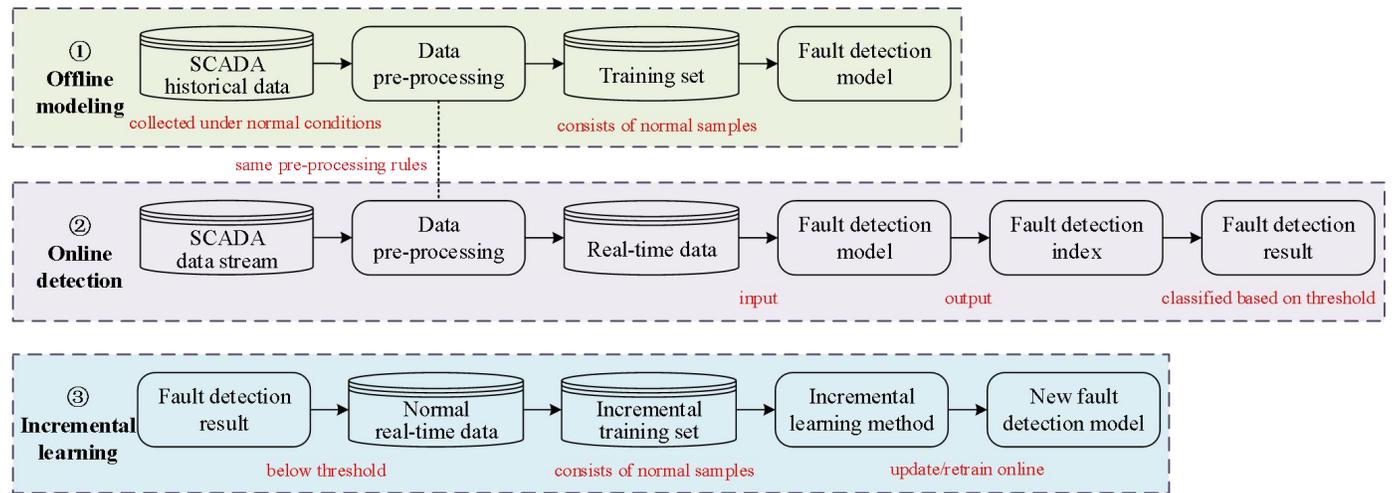


Fig. 1. Framework of WT FD models with IL.

3) Incremental Learning. Firstly, the normal real-time data that falls below the threshold constitutes an incremental training set. Secondly, based on the designed IL method, the FD model is updated or retrained online using the incremental training set, which can adapt to TVOC during long-term operation. Thirdly, the new FD model replaces the previous model and is used in Online detection.

### III. PROPOSED INCREMENTAL LEARNING METHOD CONSIDERING FALSE NEGATIVES

#### A. Error accumulation caused by false negatives

The original definition of error accumulation is the errors generated in each integration operation in numerical solution or in each component of a system are continuously accumulated and propagated, resulting in the total error that exceeds the acceptable limit [27]. Error accumulation commonly exists in various tasks, including numerical simulation [27], image processing [28], inertia navigation [29], etc. In different tasks, the forms of error accumulation are usually different. For the FD models with IL, the error accumulation is caused by false negatives, which are defined as fault samples that are misclassified as normal. The mechanism of the error accumulation is shown in Fig. 2.

As shown in Fig. 2, firstly, the FD model will inevitably output some false negatives during online detection. Secondly, false negatives will be added to the incremental training set if they are not processed. There will be more fault data in the incremental training set that should only consist of normal data, and it is equivalent to more label noise [30]. Thirdly, the low-quality incremental training set will be used to update the FD model and its accuracy will be negatively affected. Specifically, considering the correlation between different fault data, the FD model will misclassify more fault data after updating, thereby reducing the FDR and outputting more false negatives. Thus, a positive feedback mechanism driven by false negatives is built, which leads to error accumulation. The above error accumulation problem can seriously weaken the performance improvement brought about by IL and it is necessary to alleviate its negative effects.

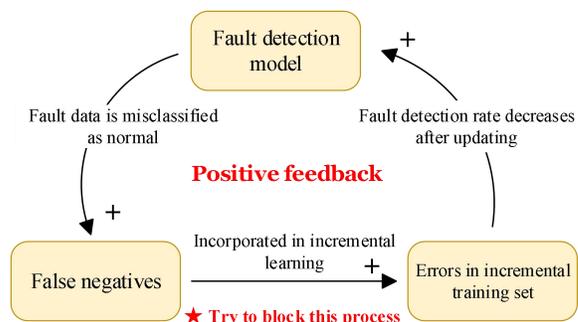


Fig. 2. Error accumulation caused by false negatives.

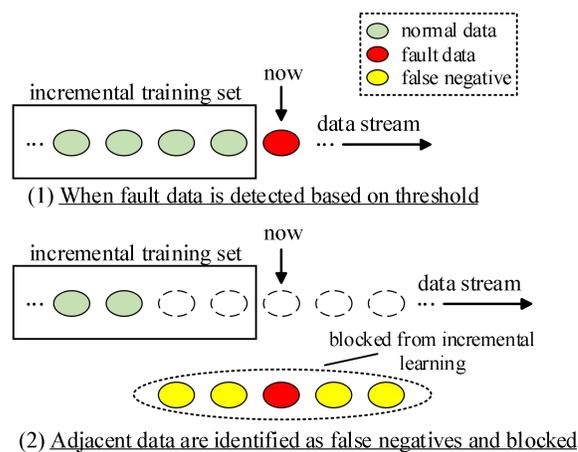


Fig. 3. Schematic diagram of proposed processing strategy for false negatives.

These false negatives in Fig. 2 are related to the strong temporal correlation in the SCADA data stream and the threshold-based hard classification strategy. Because of the strong temporal correlation, some data adjacent to the fault data have a high risk of being fault data. Because of the hard classification strategy, if the FD indexes of some data adjacent to the fault data are slightly below the FD threshold, these high-risk data will be classified as normal data and are likely to become false negatives.

Based on the above analysis, a processing strategy for false negatives is proposed as shown in Fig. 3. Firstly, the normal data in the data stream is continuously added to the incremental training set until the real-time data is classified as fault data, that is, its FD index is above the threshold. Secondly, a specific amount of data adjacent to the fault data is identified as false negatives, including previously collected data and subsequent data to be collected. These false negatives and the fault data are blocked from being used in model updating, which can reduce the errors in the incremental training set and improve the FDR of the FD model.

### B. Proposed incremental learning method

For WT FD models, a novel IL method is proposed to adaptively update the models online. The two key points of the proposed method are a data buffer that caches some new normal data used in model updating, and a processing strategy for false negatives that blocks false negatives from being used in model updating, respectively. The detailed process of the proposed IL method is shown in Fig. 4.

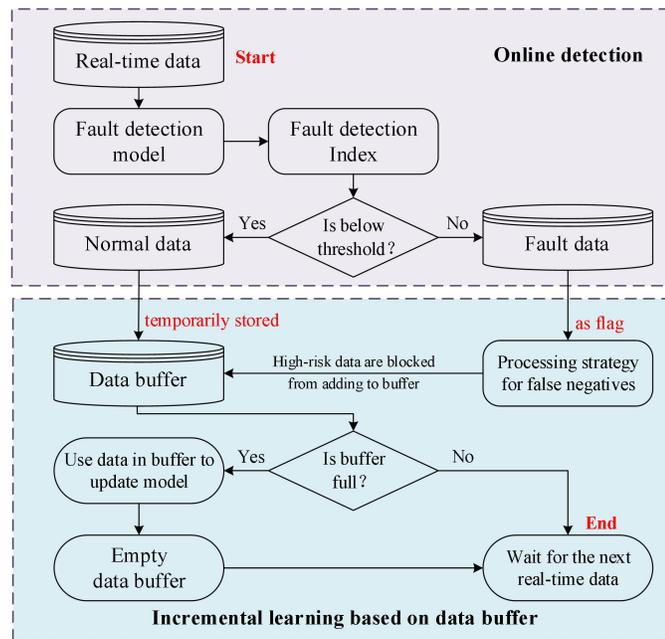


Fig. 4. Proposed IL method for WT FD models.

As shown in Fig. 4, in each computing cycle, the FD model outputs the FD index of the real-time data, and the data is classified as normal data or fault data based on the FD threshold. Then, the proposed IL method will start. Firstly, if the real-time data is fault data, it will not be added to the buffer and the processing strategy will start to delete false negatives in the buffer and block some subsequent false negatives. If the real-time data is normal, it will be added to the buffer when it is not blocked. Secondly, if the data buffer is full, the data in it will be used to update the FD model, and the buffer will be emptied after updating. If the buffer is not full, the FD model will not be updated.

One of the hyperparameters in the proposed method is the capacity of the data buffer, which determines the shortest

period for model updating. A lower capacity means a higher update frequency, and this has the advantage of higher accuracy because the model can adapt to TVOC faster. However, a higher update frequency also means higher computational costs, and the model may face more serious error accumulation. Another hyperparameter in the proposed method is the identification scope of false negatives. The larger the scope, the more data will be identified as false negatives, resulting in stronger suppression of error accumulation, but it will also increase information loss.

The proposed IL method has concise procedures and is compatible with different FD algorithms, which has good potential for practical applications.

### C. Algorithm of processing strategy for false negatives

The pseudocode of the proposed processing strategy for false negatives is shown in Algorithm. 1. One of the keys to the algorithm is to define the data structure of the buffer as a stack, that is, inserting and deleting data only at the end of the buffer. So data that are not adjacent to fault data will not be identified as false negatives. Another key is to use a pointer  $p$  to control the modification of the buffer and the flexibility of the strategy is improved.

#### Algorithm. 1. Proposed processing strategy for false negatives

**Initialization:** Data buffer (stack)  $\mathbf{D} = \{\emptyset\}$ ; Pointer for modifying buffer  $p(t) = 0$ . (In addition to running the algorithm for the first time, initialization is required when “Empty data buffer” is executed as shown in Fig. 4.)

**Input:** Real-time data  $\mathbf{x}(t)$ ; Fault flag of  $\mathbf{x}(t)$   $a(t)$ ; Data buffer  $\mathbf{D} = \{\mathbf{x}_i\}_{i=1}^m$ ; Pointer  $p(t)$ ; Identification scope of false negatives  $N$ .

**Output:** Data buffer after modifying  $\mathbf{D}$ ; Pointer  $p(t+1)$ .

{ If  $a(t)$  is true. //Real-time data is fault data.

If  $p(t) \geq N$ . //Delete false negatives previously stored.

Delete samples in  $\mathbf{D}$  from  $\mathbf{x}_m$  to  $\mathbf{x}_{m-N+1}$ .

End

If  $p(t) > 0$  and  $p(t) < N$ .

Delete samples in  $\mathbf{D}$  from  $\mathbf{x}_m$  to  $\mathbf{x}_{m-p(t)+1}$ .

End

$p(t+1) = -N$ . //Block subsequent false negatives.

Else //Real-time data is normal.

If  $p(t) \geq 0$  //Subsequent data are not blocked.

Add  $\mathbf{x}(t)$  to  $\mathbf{D}$  as  $\mathbf{x}_{m+1}$ .

End

$p(t+1) = p(t) + 1$ . //Subsequent data are blocked.

End }

In [26], a punishment mechanism is proposed for fault data and their adjacent data. However, in this mechanism, all of the data in the buffer will be deleted when the real-time data is fault data, instead of utilizing the stack and pointer in the proposed strategy. Therefore, the mechanism in [26] is not suitable for FD models with low update frequencies, as it is challenging for a large buffer to become full under TVOC. In comparison to the previous method, the proposed strategy can be applied to data buffers of various sizes and update frequencies.

IV. SCADA DATASET AND PROCESSING

A. Data description and parameter selection

The SCADA dataset used in this paper is from an onshore WT and it records overheating faults of the gearbox. The information about the SCADA dataset and the gearbox faults is shown in Tab. I.

TABLE I  
INFORMATION OF SCADA DATASET USED IN EXPERIMENTS

Type	Onshore WT	Rated power	1.5 MW
Location	Hebei province, China	Rated wind speed	11.5 m/s
Generator	Doubly fed induction generator	Cut-in wind speed	3 m/s
Sampling period of data	1 min	Cut-off wind speed	25 m/s
Time range of dataset		02-21 to 11-17, in the year 2017	
Fault records about the gearbox fault		Gearbox oil temp exceeded the upper limit (starts at 11-17)	
Fault start time estimated by professional		10-28	

TABLE II  
OPERATIONAL PARAMETERS USED IN FD EXPERIMENTS

No.	Name of parameter	Unit
P1	Main shaft speed	r/min
P2	Nacelle temp	°C
P3	Gearbox oil inlet pressure	bar
P4	Gearbox oil filter-front pressure	bar
P5	Gearbox drive bearing temp	°C
P6	Gearbox no-drive bearing temp	°C
P7	Gearbox oil temp	°C

As shown in Tab. I, it is worth noting that the auto-recorded fault start time 11-17 is later than the estimated fault start time 10-28 by professional of the wind farm. This is common in practice because conservative alarm rules and thresholds are often used in SCADA systems, resulting in incipient faults being hardly detected under TVOC [31].

To detect gearbox overheating faults and reduce the complexity of FD models, seven parameters are selected from the parameters in SCADA data, as shown in Tab. II. Main

shaft speed (P1) can reflect the load of the gearbox, which is related to the heat generation of the gearbox. Nacelle temp (P2) can reflect the external temperature of the gearbox, which is related to the heat dissipation of the gearbox. P1 and P2 are two important external factors of the TVOC of the gearbox. P3~P7 are the only five parameters that are directly acquired from the gearbox, which reflect the operating conditions of the gearbox.

B. Data pre-processing

The raw data is pre-processed using the following rules.

- 1) Delete the data that contains null values or null timestamps.
- 2) Delete the data where the active power is less than or equal to zero, and the data where the wind speed is less than the cut-in speed or greater than the cut-out speed. This is because they are collected when the WT cannot output electricity.
- 3) The three-sigma criterion is used to detect gross errors and the detected data will be deleted.
- 4) The Z-Score method is used to normalize the remaining data, which can avoid the negative effect of dimensions in different parameters.

Since approximately 19% of raw SCADA data does not contain timestamps or measured values, deletion is used in pre-processing instead of interpolation to ensure accuracy at the cost of data volume. Taking parameter P1 as an example, the results before and after pre-processing are shown in Fig. 5. Except for null values, a significant amount of raw data contains extremely low values before pre-processing, which are collected when WT was not generating electricity. After pre-processing, these outliers are deleted, reducing the interference to the FD task.

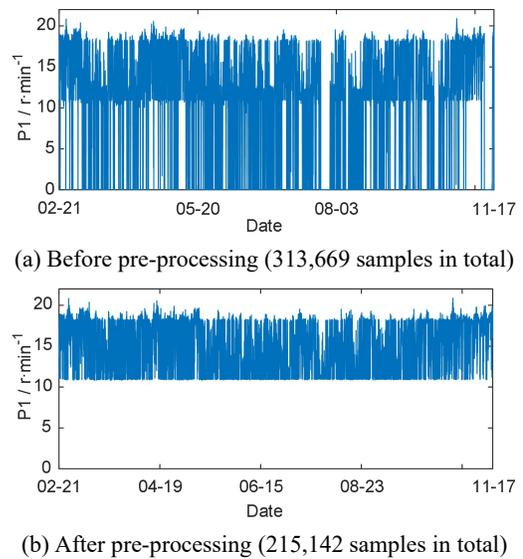


Fig. 5. Parameter P1 before and after data pre-processing.

All seven parameters after pre-processing are shown in Fig. 6. As shown in the figure, different parameters show various types of TVOC. For P1, P5 and P6, their long-term trends do not change significantly. For P2, it shows significant periodic

TVOC due to seasonal variations. For P3, it gradually decreases at the beginning and remains stable thereafter, and it decreases a lot at the final stage, which is highly relevant to the gearbox fault. For P4, its trend is similar to P3 until 08-09. From 08-09 to 08-23, it shows a prominent change that the normal range of the pressure decreases a lot. For P7, it remains stable until it increases sharply after 10-28, which is consistent with the fault description shown in Tab. I. The above various types of TVOC will lower the accuracy of WT FD models. Additionally, it is worth noting that a large amount of noise in the SCADA data is preserved after processing, which is similar to the practical scenarios of the WT FD models.

segmentation point between the normal stage and the fault stage is 10-28, which is based on professional opinions as shown in Tab. I.

TABLE III  
PARTITIONING OF DIFFERENT DATASETS

Dataset	Time span	Number of samples
Training set	02-21 to 03-31	30,434
Test set (normal stage)	04-01 to 10-27	163,608
Test set (fault stage)	10-28 to 11-17	21,100

TABLE IV  
DISTRIBUTIONS OF PARAMETERS IN DIFFERENT STAGES

Variable	Unit	Test set (normal stage)		Test set (fault stage)	
		Mean value	Standard deviation	Mean value	Standard deviation
P1	r/min	14.37	2.92	16.99	2.09
P2	°C	24.73	6.68	10.81	7.39
P3	bar	3.03	0.13	2.84	0.13
P4	bar	4.80	0.37	4.66	0.35
P5	°C	67.24	4.17	72.66	4.22
P6	°C	62.67	3.56	68.11	4.53
P7	°C	58.30	2.00	63.35	3.75

After the data segmentation, the distributions of the seven parameters in the normal stage and the fault stage are shown in Tab. IV. As shown in the table, there are significant differences in the distributions between the normal stage and fault stage, especially for the three temperature parameters P5~P7 acquired from the gearbox. Compared to the normal stage, the mean values of P5~P7 in the fault stage increase by about 5 °C, which is consistent with the overheating fault. The differences in the distributions can prove the rationality of the data segmentation point provided by the professional.

## V. EXPERIMENTS AND ANALYSIS

### A. FD models used in experiments

Four typical FD algorithms are used to verify the proposed method, and they are based on different principles as follows.

1) Gaussian kernel regression (GKR) is one of the typical multiple-input single-output parametric regression algorithms, which have been used in some FD tasks [32].

2) Multivariate state estimation technique (MSET) is one of the typical multiple-input multiple-output non-parametric regression algorithms, and MSET has been used in many FD tasks such as air compressors [33], WT [34], industrial fans [35], and aircraft [36].

3) FD tasks can be considered as a one-class classification problem, and one-class support vector machine (OCSVM) is one of the typical one-class classification algorithms [37].

4) Multivariate statistics process monitoring algorithms are

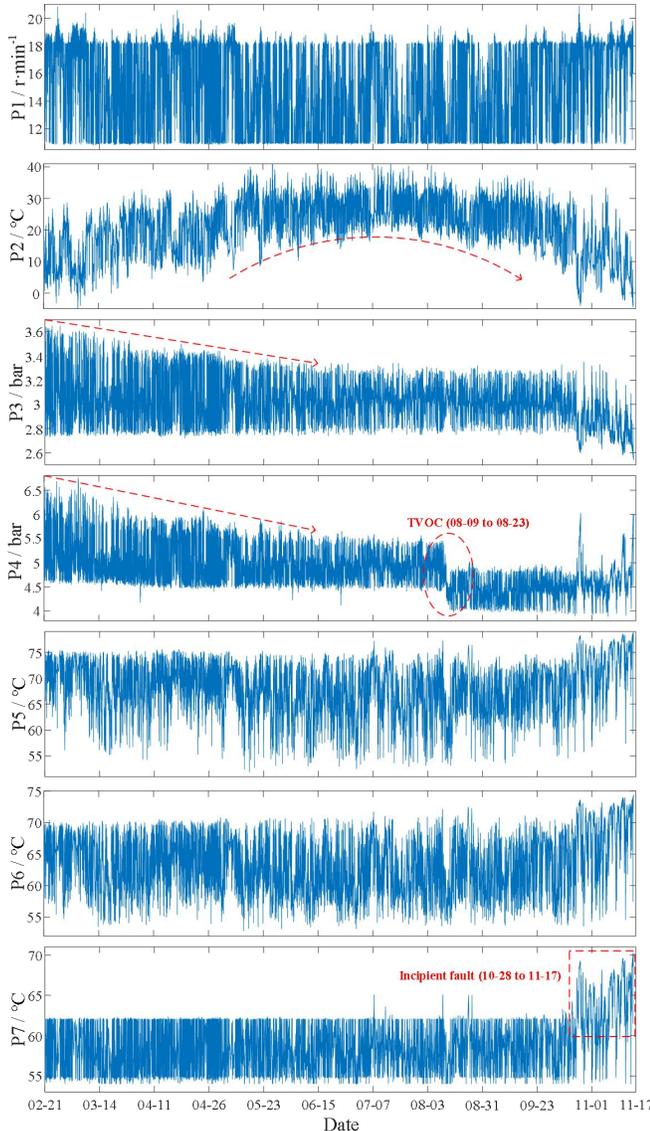


Fig. 6. Seven parameters after data pre-processing.

### C. Data segmentation

The data is divided into different datasets after pre-processing as shown in Tab. III. The training set is used to offline train the FD models. The test set is divided into the normal stage and the fault stage, and they are used to evaluate the FAR and FDR of FD models, respectively. The

widely used in FD tasks and principal component analysis (PCA) is one of the typical algorithms [38].

### (1) Gaussian kernel regression (GKR)

The function `@fitrkernel` in MATLAB is used to train the GKR model. The independent variables of the GKR model are P1~P6, and the dependent variable is P7. On the test set, the function `@predict` is used to calculate the predicted values of the dependent variable.

Residuals between the measured values and predicted values of the dependent variable P7 are used as the FD index of GKR and are defined as follows.

$$E(i) = y(i) - \hat{y}(i) \quad (1)$$

where  $E$  is the residual,  $y$  is the measured value of the dependent variable and  $\hat{y}$  is the predicted value. When the absolute value of residual  $|E|$  is greater than the designed threshold, the data is classified as fault data.

GKR's model updating strategy is incremental updating. Specifically, the function `@incrementalLearner` is used to convert the GKR model to an incremental learner defined in MATLAB and `@updateMetricsAndFit` is used to update the parameters in the model based on the incremental training data. The hyperparameters of GKR models are set to the default settings of the related functions in MATLAB.

### (2) Multivariate state estimation technique (MSET)

The essential part of MSET is the memory matrix  $D$ , which is defined as follows.

$$D = [X_1 \ X_2 \ \dots \ X_m] = \begin{bmatrix} x_1(1) & x_1(2) & \dots & x_1(m) \\ x_2(1) & x_2(2) & \dots & x_2(m) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(1) & x_n(2) & \dots & x_n(m) \end{bmatrix} \quad (2)$$

where  $X_i$  is column vector constructed of training data,  $m$  is the number of training data,  $n$  is number of parameters and all seven parameters are used. The redundant data in  $D$  will weaken the computation speed and accuracy of MSET. In this paper, a processing method for the redundant data in MSET proposed in [19] is used.

The observed vector  $X_{obs}$  and the estimated vector  $X_{est}$  are the input and output of MSET, respectively.  $X_{obs}$  is the real-time data and  $X_{est}$  is the estimated result of  $X_{obs}$ . The estimated vector  $X_{est}$  is calculated as follows.

$$X_{est} = D \cdot (D^T \otimes D)^{-1} \cdot D^T \otimes X_{obs} \quad (3)$$

where  $\otimes$  is a nonlinear operator and Euclidean metric is used in this paper.

According to the similarity principle, if  $X_{obs}$  is normal, there will be high similarities between  $X_{obs}$  and some vectors in  $D$ , which leads to the small difference between  $X_{obs}$  and  $X_{est}$ . By contrast, the difference will be larger if  $X_{obs}$  is fault data. The square prediction error (SPE) between  $X_{obs}$  and  $X_{est}$  is used as the FD index of MSET and is defined as follows.

$$SPE(i) = \sum_{j=1}^n [X_{est}^j(i) - X_{obs}^j(i)]^2 \quad (4)$$

where  $X_{est}^j$  and  $X_{obs}^j$  are the elements in vector  $X_{est}$  and  $X_{obs}$ , respectively. When SPE is greater than the designed threshold, the data is classified as fault data.

MSET's model updating strategy is incremental updating. Specifically, the incremental data is added into  $D$ , and the sample selection method is used to delete the redundant data. The hyperparameter in MSET models is the threshold of the redundant data and it takes 0.3 in this paper.

### (3) One-class support vector machine (OCSVM)

The function `@fitsvm` in MATLAB is used to train the OCSVM model and all seven parameters are used. On the test set, the function `@predict` is used to calculate the score of OCSVM. If score  $\geq 0$ , the data is classified as the positive class (normal), and if score  $< 0$ , it is classified as the negative class (fault). The value of score indicates the confidence level of the classification result. In this paper, the FD index score<sub>FD</sub> of OCSVM is defined as follows.

$$\text{score}_{\text{FD}}(i) = \begin{cases} 0, & \text{score}(i) \geq 0 \\ -\text{score}(i), & \text{score}(i) < 0 \end{cases} \quad (5)$$

where only the negative class is considered in the FD index and greater score<sub>FD</sub>( $i$ ) means higher probability of being fault data. When score<sub>FD</sub> is greater than the designed threshold, the data is classified as fault data.

OCSVM's model updating strategy is retraining the entire model online. Firstly, the entire training set is stored, and the incremental data is added to the training set when the buffer is full. Secondly, the new training set and `@fitsvm` are used to train a new OCSVM model. The hyperparameters of OCSVM models are set to the default settings of the related functions in MATLAB.

### (4) Principal component analysis (PCA)

For the training of PCA model, firstly, the training data is constructed as the matrix  $X_{tr}$  and the covariance matrix  $R$  of  $X_{tr}$  is calculated. Secondly, calculate the eigenvalues  $\lambda_i$  and eigenvectors  $P_i$  of covariance matrix  $R$ , and arrange  $\lambda_i$  and  $P_i$  in descending order. Thirdly,  $P_i$  of the first  $k$  eigenvalues are used to construct the load matrix  $P = [P_1, P_2, \dots, P_k]$ . In this paper, hyperparameter of PCA  $k = 3$ .

On the test set, the reconstructed vector  $X_{re}$  of the input vector  $X_{in}$  is calculated as follows.

$$X_{re} = X_{in} P P^T \quad (6)$$

Like MSET in (4), SPE between  $X_{in}$  and  $X_{re}$  is used as the FD index of PCA.

PCA's model updating strategy is retraining the entire model like OCSVM. Specifically, the new training set is used to calculated a new load matrix  $P$ .

## B. Comparative experiments

On the test set, the FD results of different FD algorithms are shown in Fig. 7 to Fig. 10. The black dashed lines in the figures are used to distinguish between normal and fault stages. The red lines represent the FD thresholds. For fault alarms,

considering the noise in SCADA data, if 30 consecutive FD indexes exceed the threshold, it will issue a fault alarm.

As shown in Fig. 7 and Fig. 8, for both GKR and MSET, after using the proposed IL method, the FD indexes become more stable and there are fewer fault alarms in the normal stage. When false negatives are not processed, although there are fewer false alarms, the alarms in the fault stage also decrease. As shown in Fig. 9, for OCSVM, the amplitude and sensitivity of the FD indexes are improved after using the IL method, which leads to fewer false alarms. As shown in Fig. 10, the FD results show little difference between the three PCA models. As shown in the results of the fault alarms, MSET and OCSVM models have better robustness to noise, as they send fewer false alarms. On the other hand, GKR and PCA models have poorer robustness to noise.

Following indicators are used to quantitatively evaluate the performance of different FD models.  $FAR$  is used to evaluate the performance in the normal stage and  $FDR$  is used to evaluate the performance in the fault stage.  $FAR$  and  $FDR$  are calculated as follows.

$$FAR = FP/m_1 \quad (7)$$

$$FDR = TP/m_2 \quad (8)$$

where  $m_1$  and  $m_2$  are the number of samples in the normal and fault stage of test set, respectively. As shown in Tab. III,  $m_1 = 30,434$  and  $m_2 = 184,708$ .  $FP$  is the number of false positives that normal samples are misclassified as fault samples, and  $TP$  is the number of true positives that fault samples are correctly classified as fault samples.

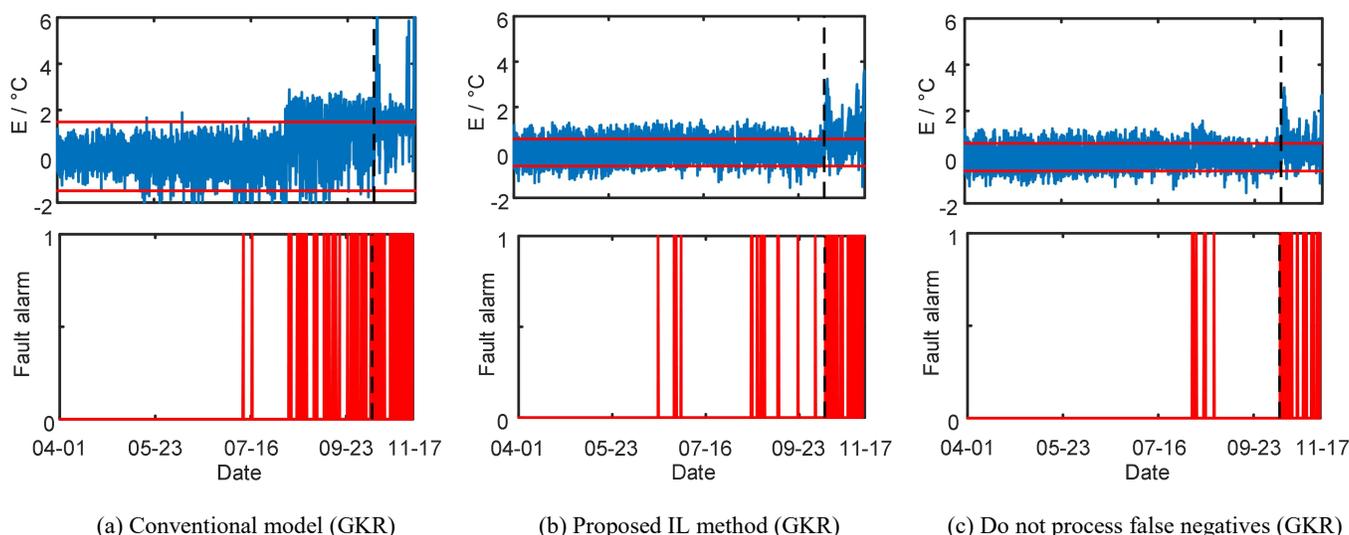


Fig. 7. Fault detection results of different GKR models.

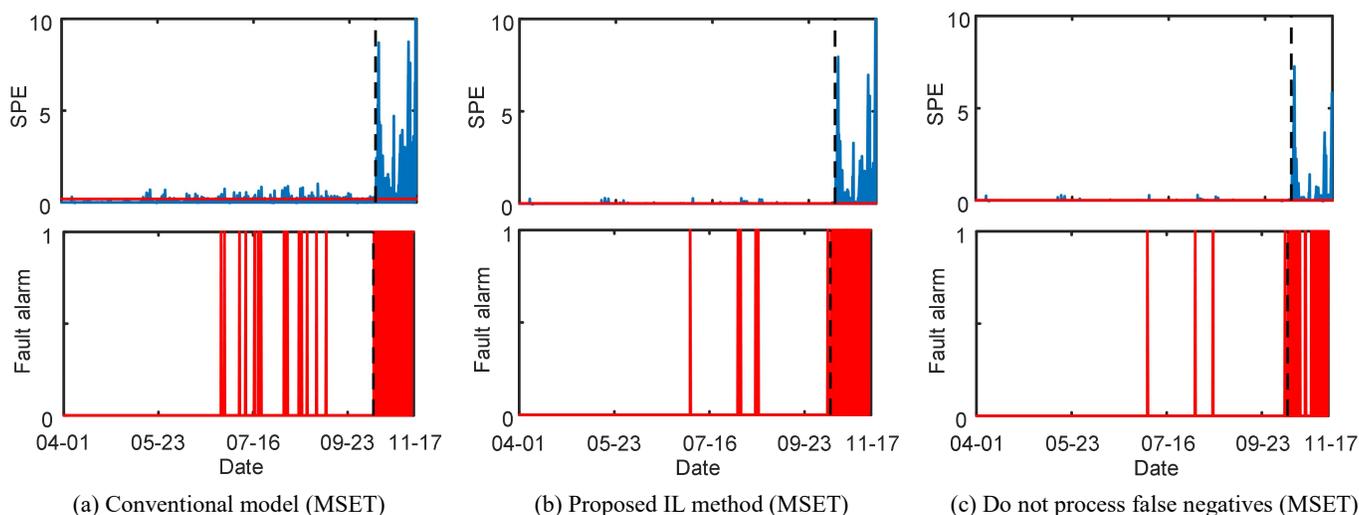


Fig. 8. Fault detection results of different MSET models.

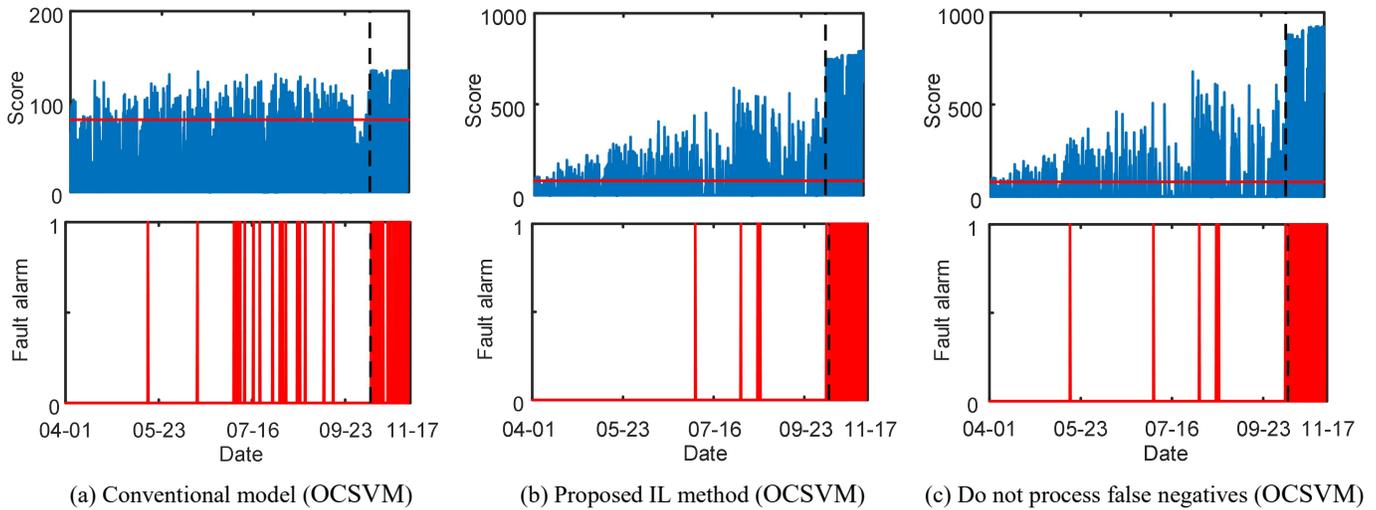


Fig. 9. Fault detection results of different OCSVM models.

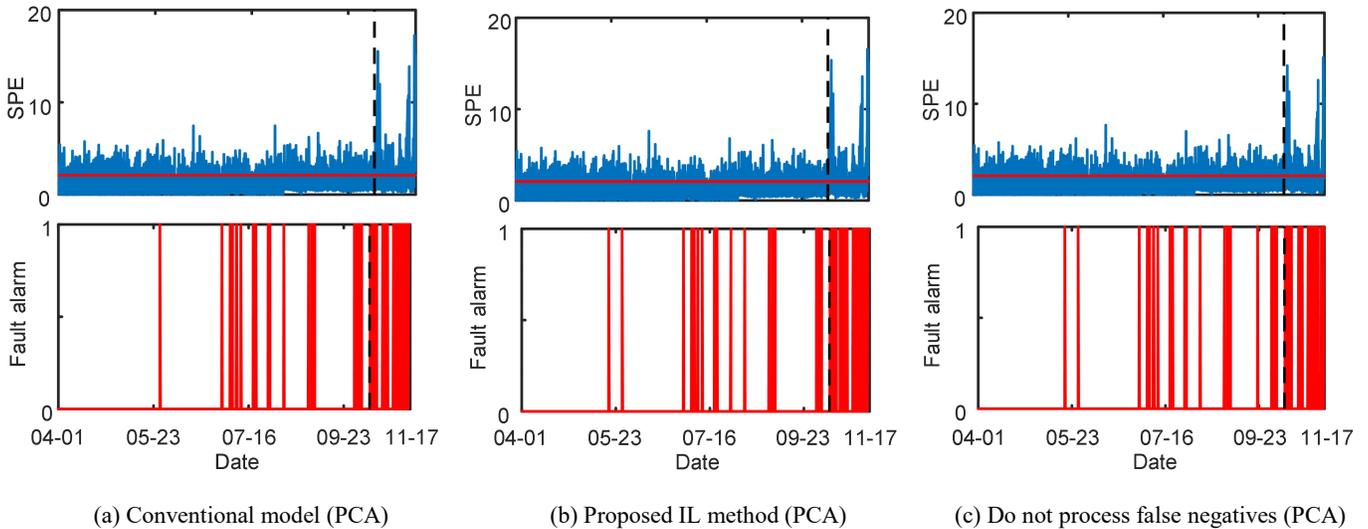


Fig. 10. Fault detection results of different PCA models.

TABLE V  
COMPARATIVE EXPERIMENTAL RESULTS OF DIFFERENT FD MODELS

Algorithm	IL method	Threshold	$C$	$N$	$FAR$	$FDR$	$T$	Calculation time on test set
GKR	None	1.5	—	—	6.66%	43.18%	—	0.6 s
	Proposed method	0.6	6,000	20	4.16%	43.18%	20	13.5 s
	Do not process false negatives	0.6	6,000	—	3.59%	32.63%	28	14.5 s
MSET	None	0.19	—	—	1.48%	45.26%	—	4.5 s
	Proposed method	0.02	200	30	0.46%	45.26%	844	3058.1 s
	Do not process false negatives	0.02	200	—	0.35%	31.24%	887	3445.4 s
OCSVM	None	80	—	—	2.04%	41.14%	—	5.5 s
	Proposed method	80	5,000	30	1.20%	48.79%	28	4457.6 s
	Do not process false negatives	80	5,000	—	0.85%	46.18%	34	6462.1 s
PCA	None	2.1	—	—	6.41%	41.61%	—	0.5 s
	Proposed method	2.1	5,000	110	6.20%	43.18%	7	1.0 s
	Do not process false negatives	2.1	5,000	—	6.02%	35.11%	33	1.8 s

The evaluation indicators of different FD models are shown in Tab. V. In the table,  $C$  is the capacity of the data buffer and  $N$  is the identification scope of false negatives.  $T$  is the updating times of FD models and is used to evaluate the updating frequency. For the GKR models, the median of these indicators in 500 repeated experiments are shown. The runtime of the experiments is MATLAB 2023b, i7-10700 CPU and 16GB RAM.

As shown in Tab. V, for GKR and MSET, models with IL have lower  $FAR$  than the conventional models when they have the same  $FDR$ . For OCSVM and PCA, the IL method can improve both the  $FAR$  and  $FDR$ . If false negatives are not processed, the  $FAR$  of GKR, MSET, OCSVM, and PCA slightly decrease by 0.57% (4.16%-3.59%), 0.11% (0.46%-0.35%), 0.35% (1.20%-0.85%), and 0.18% (6.20%-6.02%), respectively. However, the  $FDR$  of the four models decrease a lot by 10.55% (43.18%-32.63%), 14.02% (45.26%-31.24%), 2.61% (48.79%-46.18%), and 8.07% (43.18%-35.11%), respectively, which proves that the processing of false negatives has a significant improvement on  $FDR$ . Under the noise shown in Fig. 6, MSET and OCSVM models show better robustness, as their  $FAR$  is below 2.1%. On the other hand, GKR and PCA models show poorer robustness, as their  $FAR$  is above 3.5%.

For the computational complexity, when the IL method is not used, the calculation time of the four models is short because model updating is not needed. When using the IL method, the calculation time of GKR and PCA models is still short, because of the low update frequency and the fast incremental updating or re-training speed. On the other hand, the calculation time of MSET and OCSVM models with IL increases a lot, because the update frequency of MSET is high and the re-training speed of OCSVM is slow. Besides the effect on  $FAR$  and  $FDR$ , processing false negatives can lower the updating frequency of models to lower the calculation time.

Besides the real-time curves of FD indexes and fault alarms, daily reports of fault alarms are often used in practical applications, which record the  $FDR$  within one day. For MSET models, the daily reports of fault alarms in the fault stage of the test set are shown in Fig. 11. In the figure, there are no records from 11-13 to 11-15 because less than 10% of SCADA data remains after data pre-processing.

As shown in Fig. 11, for the MSET model using the IL method, the  $FDR$  at 11-07, 11-10, 11-12 and 11-17 exceed 75%, which means the gearbox fault can be detected at least 5 to 10 days earlier than the recorded fault start time 11-17. However, when false negatives are not processed,  $FDR$  decreases a lot on all days.

### C. FD threshold selection

For the conventional FD models, their FD thresholds are selected based on the receiver operating characteristic (ROC) curves, where the horizontal axis is the  $FAR$  on the test set and the vertical axis is the  $FDR$ . The ROC curves used are drawn based on the variations of the  $FAR$  and  $FDR$  with different FD thresholds. For example, the ROC curves of MSET and OCSVM models are shown in Fig. 12.

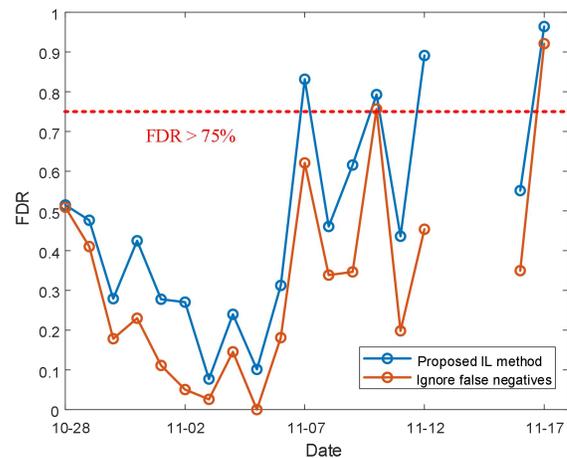
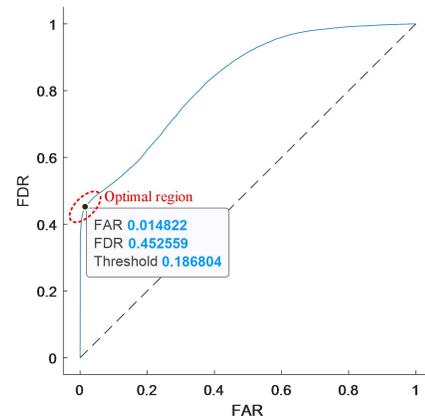
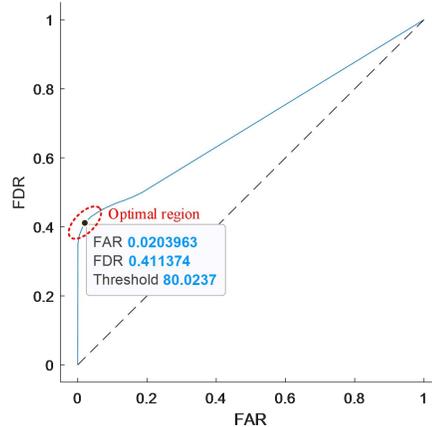


Fig. 11. Daily reports of fault alarms for MSET models.



(a) ROC curve of MSET model



(b) ROC curve of OCSVM model

Fig. 12. ROC curves and FD threshold selections of MSET model and OCSVM model.

According to the properties of the ROC curve, the better the model performance, the closer the points on the curve are to (0,1). Additionally, considering the difficulty in accepting high  $FAR$  in practical applications, the optimal regions shown in Fig. 12 are determined. The FD thresholds are selected in the optimal regions, with the threshold of MSET being 0.19 and the threshold of OCSVM being 80, which leads to lower  $FAR$  and acceptable  $FDR$ .

IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT

For the FD models with IL, firstly thresholds used in the conventional models are selected, and the *FAR* and *FDR* of the IL models are calculated. Compared to the conventional model, if the *FAR* of the IL model decreases and the *FDR* increases, the threshold used in the conventional model still will be selected for the IL model, like OCSVM and PCA. If the *FAR* and *FDR* of the IL model both decrease compared to the conventional model, the threshold will be adjusted until the IL model and the conventional model have the same *FDR* to compare the performance on the same baseline, like GKR and MSET.

D. Hyperparameter experiments

There are two hyperparameters in the proposed IL method: the capacity of the data buffer *C* and the identification scope of false negatives *N*. The two hyperparameters can affect the performance of FD models in different ways. Therefore, hyperparameter experiments and sensitivity analysis are performed on two different FD algorithms, GKR and MSET. The experimental results on GKR models are shown in Tab. VI and Tab. VII.

TABLE VI  
EXPERIMENTAL RESULTS OF HYPERPARAMETER *C* ON GKR

<i>C</i>	<i>FAR</i>	<i>FDR</i>	<i>T</i>
15,000	6.75%	53.03%	7
12,000	4.42%	43.03%	10
8,000	4.23%	42.35%	15
6,000 (baseline)	4.16%	43.18%	20
5,000	4.20%	42.01%	25
4,000	4.18%	38.07%	31
2,000	4.27%	24.77%	64

TABLE VII  
EXPERIMENTAL RESULTS OF HYPERPARAMETER *N* ON GKR

<i>N</i>	<i>FAR</i>	<i>FDR</i>	<i>T</i>
80	6.48%	51.12%	8
60	5.16%	43.98%	12
40	4.73%	41.85%	16
30	4.33%	42.49%	18
20 (baseline)	4.16%	43.18%	20
10	3.61%	41.11%	24
5	3.86%	33.48%	25

As shown in Tab. VI, on the whole, greater *C* means a lower update frequency, higher *FAR*, and higher *FDR*. And as shown in Tab. VII, greater *N* means more data will be classified as false negatives during updating, which leads to a lower update frequency, higher *FAR*, and higher *FDR*.

For the sensitivity analysis of GKR, when *C* is within the range of 5,000 to 12,000, *FAR* is within the range of 4.16% to 4.42% and *FDR* is within the range of 42.01% to 43.18%, which are relatively stable. When *N* is within the range of 10 to 40, *FAR* is within the range of 3.61% to 4.73% and *FDR* is within the range of 41.11% to 43.18%, which are relatively stable.

The experimental results on MSET models are shown in Tab. VIII and Tab. IX.

TABLE VIII  
EXPERIMENTAL RESULTS OF HYPERPARAMETER *C* ON MSET

<i>C</i>	<i>FAR</i>	<i>FDR</i>	<i>T</i>
5,000	0.77%	48.09%	32
2,000	0.56%	46.45%	83
1,000	0.51%	45.81%	168
400	0.49%	44.15%	422
200 (baseline)	0.46%	45.26%	844
100	0.43%	44.12%	1696
50	0.38%	44.67%	3400

TABLE IX  
EXPERIMENTAL RESULTS OF HYPERPARAMETER *N* ON MSET

<i>N</i>	<i>FAR</i>	<i>FDR</i>	<i>T</i>
80	0.60%	47.84%	797
60	0.56%	45.79%	811
40	0.48%	45.21%	836
30 (baseline)	0.46%	45.26%	844
20	0.48%	44.60%	852
10	0.44%	41.71%	865
5	0.39%	40.24%	871

As shown in Tab. VIII, compared with GKR, the effect of *C* on the performance of MSET models is the same, but a higher update frequency is needed for MSET. As shown in Tab. IX, the effect of *N* on the *FAR* and *FDR* of MSET models is the same as that of GKR, but the effect on the update frequency is relatively less.

For the sensitivity analysis of MSET, when *C* is within the range of 50 to 2,000, *FAR* is within the range of 0.38% to 0.56% and *FDR* is within the range of 44.12% to 46.45%, which are relatively stable. When *N* is within the range of 20 to 80, *FAR* is within the range of 0.46% to 0.60% and *FDR* is within the range of 44.60% to 47.84%, which are relatively stable. Compared to GKR, the optimal ranges of hyperparameters are wider for MSET, which shows better adaptability in practical applications.

According to the above results, the two hyperparameters have similar effects on the performance of FD models, that is, greater values lead to a lower update frequency, higher *FAR*,

and higher FDR. The recommended selection strategy of the hyperparameters is to first choose a greater  $C$  to lower calculation time under the premise that the FAR and FDR are acceptable, and then optimize  $N$  to improve the FAR and FDR.

*E. Early stopping experiments*

In the above experiments, the FD models with IL will keep updating online until WT is shut down for maintenance due to the gearbox fault at 11-17. After the potential faults are detected, if the FD models can stop updating earlier than passive stopping, their performance will be improved because of less error accumulation caused by false negatives.

The results of early stopping experiments on different FD models are shown in Tab. X, which stops updating instantly when potential faults are detected at 10-28. In the table,  $FDR$  and  $T$  are the results without stopping updating, and  $FDR^*$  and  $T^*$  are the results with stopping updating. PCA is not included because it is not updated after 10-28.

As shown in Tab. X, stopping updating earlier can higher  $FDR$  and lower the computational costs. However, instantly stopping is infeasible in practical applications, because it may lead to unexpected stopping in the normal stage. The recommended early stopping timing is to stop updating after obtaining reliable FD results, which leads to inevitable delay compared with instantly stopping and will weaken its positive effect. For MSET, the experimental results of the different delays in early stopping are shown in Tab. XI.

As shown in Tab. XI, the longer the delay, the greater the decrease in  $FDR$ , and the result of a delay of less than two days is close to that of instantly stopping, while the result of a delay of four days is close to that of no stopping.

TABLE X  
RESULTS OF DIFFERENT FD ALGORITHMS WHEN EARLY STOPPING

Algorithm	$FDR$	$FDR^*$	$T$	$T^*$
GKR	43.18%	46.99%	20	19
MSET	45.26%	48.62%	844	799
OCSVM	48.79%	51.09%	28	20

TABLE XI  
RESULTS OF DIFFERENT DELAYS OF EARLY STOPPING FOR MSET

Delay of early stopping	$FDR$	$T$
No delay	48.62%	799
One day-delay	47.98%	800
Two days-delay	47.36%	803
Four days-delay	45.52%	811

Based on the above experimental results, for the early stopping of FD models with IL in practical applications, the recommended strategy is to use the daily reports of the FDR shown in Fig. 11. For example, if the FDR exceeds 40% for

two consecutive days, the FD models will automatically stop updating.

VI. CONCLUSION

A novel IL method is proposed to improve the performance of WT FD models under TVOC. A data buffer is built to cache some new data used in model updating. A processing strategy for false negatives is proposed to weaken the error accumulation, which leads to a continuous decrease in the FDR of FD models during online updating.

Using a real-world SCADA dataset of WT, the results show that the proposed IL method can lower the FAR of FD models and can work on different FD algorithms. Additionally, the processing of false negatives can significantly improve the FDR of FD models. The results of hyperparameter and early stopping experiments show that the proposed IL method has good potential for practical applications.

Future research will focus on optimizing the update frequency of WT FD models with IL, such as event-triggered updating, to improve computational efficiency.

REFERENCES

- [1] H. Badihi, Y. Zhang, B. Jiang, P. Pillay, and S. Rakheja, "A Comprehensive Review on Signal-Based and Model-Based Condition Monitoring of Wind Turbines: Fault Diagnosis and Lifetime Prognosis," *Proc. IEEE*, vol. 110, no. 6, pp. 754-806, 2022.
- [2] X. Jin, Z. Xu, and W. Qiao, "Condition Monitoring of Wind Turbine Generators Using SCADA Data Analysis," *IEEE Trans. Sustain. Energy*, vol. 12, no. 1, pp. 202-210, 2021.
- [3] Q. Lu, W. Ye, and L. Yin, "ResDenIncepNet-CBAM with principal component analysis for wind turbine blade cracking fault prediction with only short time scale SCADA data," *Measurement*, vol. 212, p. 112696, 2023.
- [4] L. Xiang, P. Wang, X. Yang, A. Hu, and H. Su, "Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism," *Measurement*, vol. 175, p. 109094, 2021.
- [5] Q. He, Y. Pang, G. Jiang, and P. Xie, "A Spatio-Temporal Multiscale Neural Network Approach for Wind Turbine Fault Diagnosis With Imbalanced SCADA Data," *IEEE Trans. Ind. Inform.*, vol. 17, no. 10, pp. 6875-6884, 2021.
- [6] S. Sun, W. Hu, Y. Liu, T. Wang, and F. Chu, "Matching contrastive learning: An effective and intelligent method for wind turbine fault diagnosis with imbalanced SCADA data," *Expert Syst. Appl.*, vol. 223, p. 119891, 2023.
- [7] M. Schlechtingen, I. F. Santos, and S. Achiche, "Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description," *Appl. Soft. Comput.*, vol. 13, no. 1, pp. 259-270, 2013.
- [8] M. Schlechtingen and I. F. Santos, "Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 2: Application examples," *Appl. Soft. Comput.*, vol. 14, pp. 447-460, 2014.
- [9] G. Zhang, Y. Li, W. Jiang, and L. Shu, "A fault diagnosis method for wind turbines with limited labeled data based on balanced joint adaptive network," *Neurocomputing*, vol. 481, pp. 133-153, 2022.
- [10] G. Liu, J. Si, W. Meng, Q. Yang, and C. Li, "Wind Turbine Fault Detection With Multimodule Feature Extraction Network and Adaptive Strategy," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1-13, 2023.
- [11] Y. Zhang, Y. Han, C. Wang, J. Wang, and Q. Zhao, "A dynamic threshold method for wind turbine fault detection based on spatial-temporal neural network," *J. Renew. Sustain. Energy*, vol. 14, no. 5, p. 053304, 2022.
- [12] X. Jin, H. Pan, C. Ying, Z. Kong, Z. Xu, and B. Zhang, "Condition Monitoring of Wind Turbine Generator Based on Transfer Learning and One-Class Classifier," *IEEE Sens. J.*, vol. 22, no. 24, pp. 24130-24139, 2022.
- [13] W. Liu and H. Ren, "A novel wind turbine health condition monitoring method based on common features distribution adaptation," *Int. J. Energy Res.*, vol. 44, no. 11, pp. 8681-8688, 2020.

IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT

[14]G. Jiang, W. Li, W. Fan, Q. He, and P. Xie, "TempGNN: A Temperature-Based Graph Neural Network Model for System-Level Monitoring of Wind Turbines With SCADA Data," *IEEE Sens. J.*, vol. 22, no. 23, pp. 22894-22907, 2022.

[15]X. Liu, J. Du, and Z. Ye, "A Condition Monitoring and Fault Isolation System for Wind Turbine Based on SCADA Data," *IEEE Trans. Ind. Inform.*, vol. 18, no. 2, pp. 986-995, 2022.

[16]Y. Luo, L. Yin, W. Bai, and K. Mao, "An Appraisal of Incremental Learning Methods," *Entropy*, vol. 22, no. 11, p. 1190, 2020.

[17]B. Chen, S. Yu, Y. Yu, and Y. Zhou, "Acoustical damage detection of wind turbine blade using the improved incremental support vector data description," *Renew. Energy*, vol. 156, pp. 548-557, 2020.

[18]J. Yuan, Z. Liang, R. Wang, Y. Li, Z. Wang, and J. Gao, "A novel self-learning framework for fault identification of wind turbine drive bearings," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 237, no. 7, pp. 1296-1312, 2023.

[19]Z. Wang, C. Liu, and F. Yan, "Condition monitoring of wind turbine based on incremental learning and multivariate state estimation technique," *Renew. Energy*, vol. 184, pp. 343-360, 2022.

[20]Y. Zhou, X. Tian, C. Zhang, Y. Zhao, and T. Li, "Elastic weight consolidation-based adaptive neural networks for dynamic building energy load prediction modeling," *Energy Build.*, vol. 265, p. 112098, 2022.

[21]Y. Yang, B. Chen, and H. Liu, "Bayesian compression for dynamically expandable networks," *Pattern Recognit.*, vol. 122, p. 108260, 2022.

[22]J. Zheng, H. Xiong, Y. Zhang, K. Su, and Z. Hu, "Bearing Fault Diagnosis via Incremental Learning Based on the Repeated Replay Using Memory Indexing (R-REMIND) Method," *Machines*, vol. 10, no. 5, p. 338, 2022.

[23]J. Kirkpatrick *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521-3526, 2017.

[24]Z. Yang and Z. Ge, "On Paradigm of Industrial Big Data Analytics: From Evolution to Revolution," *IEEE Trans. Ind. Inform.*, vol. 18, no. 12, pp. 8373-8388, 2022.

[25]M. Zheng, J. Man, D. Wang, Y. Chen, Q. Li, and Y. Liu, "Semi-supervised multivariate time series anomaly detection for wind turbines using generator SCADA data," *Reliab. Eng. Syst. Saf.*, vol. 235, p. 109235, 2023.

[26]Z. Wang, X. Jin, and Z. Xu, "An Adaptive Condition Monitoring Method of Wind Turbines Based on Multivariate State Estimation Technique and Continual Learning," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1-9, 2023.

[27]N. N. Smirnov, V. B. Betelin, V. F. Nikitin, L. I. Stamov, and D. I. Altoukhov, "Accumulation of errors in numerical simulations of chemically reacting gas dynamics," *Acta Astronaut.*, vol. 117, pp. 338-355, 2015.

[28]R. Eschbach and M. Pedersen, "On Large Local Error Accumulation in Multilevel Error Diffusion," *J. Imaging Sci. Technol.*, vol. 60, no. 6, pp. 060403-1-060403-9, 2016.

[29]X. Liu, Q. Zhou, X. Chen, L. Fan, and C. Cheng, "Bias-Error Accumulation Analysis for Inertial Navigation Methods," *IEEE Signal Process. Lett.*, vol. 29, pp. 299-303, 2022.

[30]M. Qian, Y. Li, and T. Han, "Positive-Unlabeled Learning-Based Hybrid Deep Network for Intelligent Fault Detection," *IEEE Trans. Ind. Inform.*, vol. 18, no. 7, pp. 4510-4519, 2022.

[31]A. Wang, Z. Qian, Y. Pei, and B. Jing, "A de-ambiguous condition monitoring scheme for wind turbines using least squares generative adversarial networks," *Renew. Energy*, vol. 185, pp. 267-279, 2022.

[32]Z. Li, S. Bao, X. Peng, and L. Luo, "Fault detection and diagnosis in multivariate systems using multiple correlation regression," *Control Eng. Practice*, vol. 116, p. 104916, 2021.

[33]C. Cui, W. Lin, Y. Yang, X. Kuang, and Y. Xiao, "A novel fault measure and early warning system for air compressor," *Measurement*, vol. 135, pp. 593-605, 2019.

[34]Z. Wang and C. Liu, "Wind turbine condition monitoring based on a novel multivariate state estimation technique," *Measurement*, vol. 168, p. 108388, 2021.

[35]Y. Lv, F. Fang, T. Yang, and C. E. Romero, "An early fault detection method for induced draft fans based on MSET with informative memory matrix selection," *ISA Trans.*, vol. 102, pp. 325-334, 2020.

[36]S. Su, Y. Sun, L. Li, C. Peng, H. Zhang, and T. Zhang, "Risk Warning for Aircraft Bleed Air System with Multivariate State Estimation Technique," *Journal of Aerospace Information Systems*, vol. 19, no. 8, pp. 550-564, 2022.

[37]C. Li, D. Cabrera, F. Sancho, M. Cerrada, R. Sánchez, and E. Estupinan, "From fault detection to one-class severity discrimination of 3D printers with one-class support vector machine," *ISA Trans.*, vol. 110, pp. 357-367, 2021.

[38]K. Zhang, B. Tang, L. Deng, and X. Yu, "Fault Detection of Wind Turbines by Subspace Reconstruction-Based Robust Kernel Principal Component Analysis," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-11, 2021.



**Ziqi Wang** received the B.Eng. Degree and Ph.D. degree in control science and engineering from North China Electric Power University, Beijing, China, in 2017 and 2022, respectively.

He is currently a Postdoctoral Researcher with the College of Control Science and Engineering, Zhejiang University, Zhejiang, China. His research interests include data analysis and data-driven modeling in wind turbine condition monitoring.



**Xiaohang Jin** (Senior Member, IEEE) received the Ph.D. degree in electronic engineering from the City University of Hong Kong, Hong Kong, China, in 2014.

Since 2014, he has been with the Zhejiang University of Technology, where he is currently an Associate Professor with the College of Mechanical Engineering. His research interests include intelligent manufacturing, condition monitoring, prognostics and health management.



**Zhengguo Xu** (Member, IEEE) received the B.S. degree in electrical engineering from Changsha Electric Power University, Changsha, China, in 2002, and the M.S. and Ph.D. degrees in control science and engineering from Tsinghua University, Beijing, China, in 2009.

He is currently a Professor with the College of Control Science and Engineering, Zhejiang University, Hang Zhou, Zhejiang, China. His research interests include data analysis and artificial intelligence applications in energy systems, fault diagnosis and prediction, and reliability analysis.